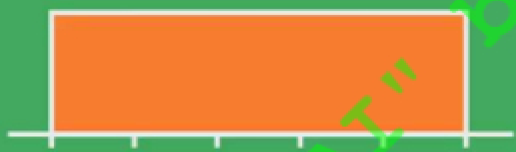
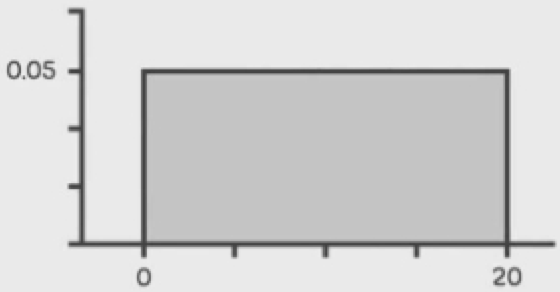
UNIFORM DISTRIBUTION



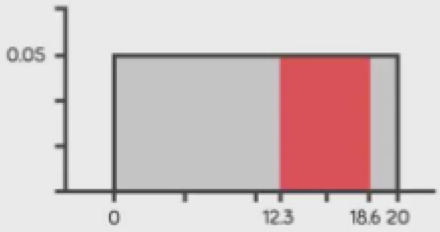
PRACTICE QUESTIONS

❸ For the uniform distribution below, what proportion of values are located between 12.3 and 18.6?



PRACTICE QUESTIONS

❸ For the uniform distribution below, what proportion of values are located between 12.3 and 18.6?



Area = L x W

= 6.3 x 0.05

Area = 0.315

Probability Distribution

Probability Distribution

Graphical representation of variable & respective probabilities of variable.

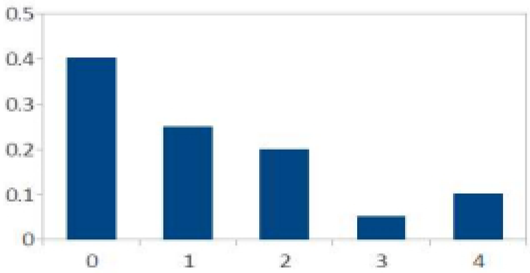
Types of Probability Distribution

- Discrete Probability Distribution
- Continuous Probability Distribution

Discrete Probability Distribution

The daily sales of large flat panel TVs at a store (X)

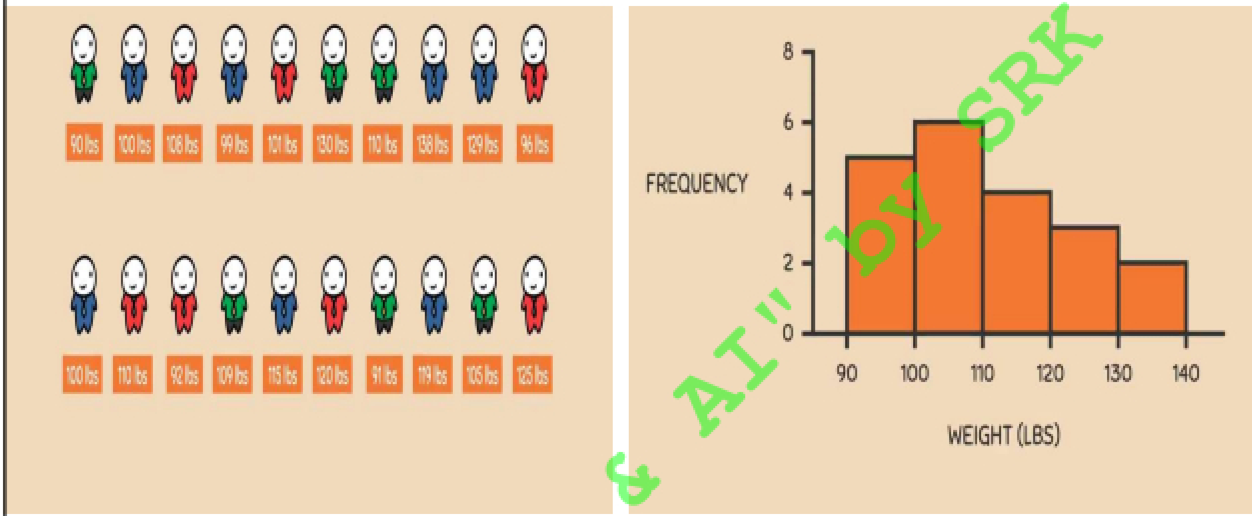
x	P(X=x)
0	0.40
1	0.25
2	0.20
3	0.05
4	0.10



What is the probability of a sale?

What is the probability of selling at least three TVs?

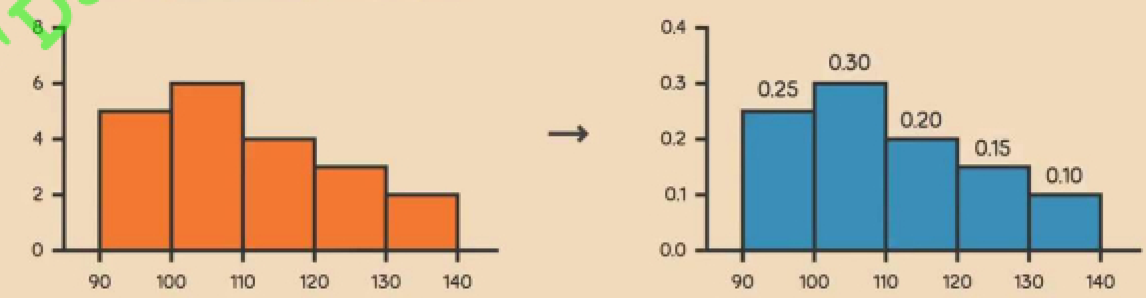
Continous Probability Distribution

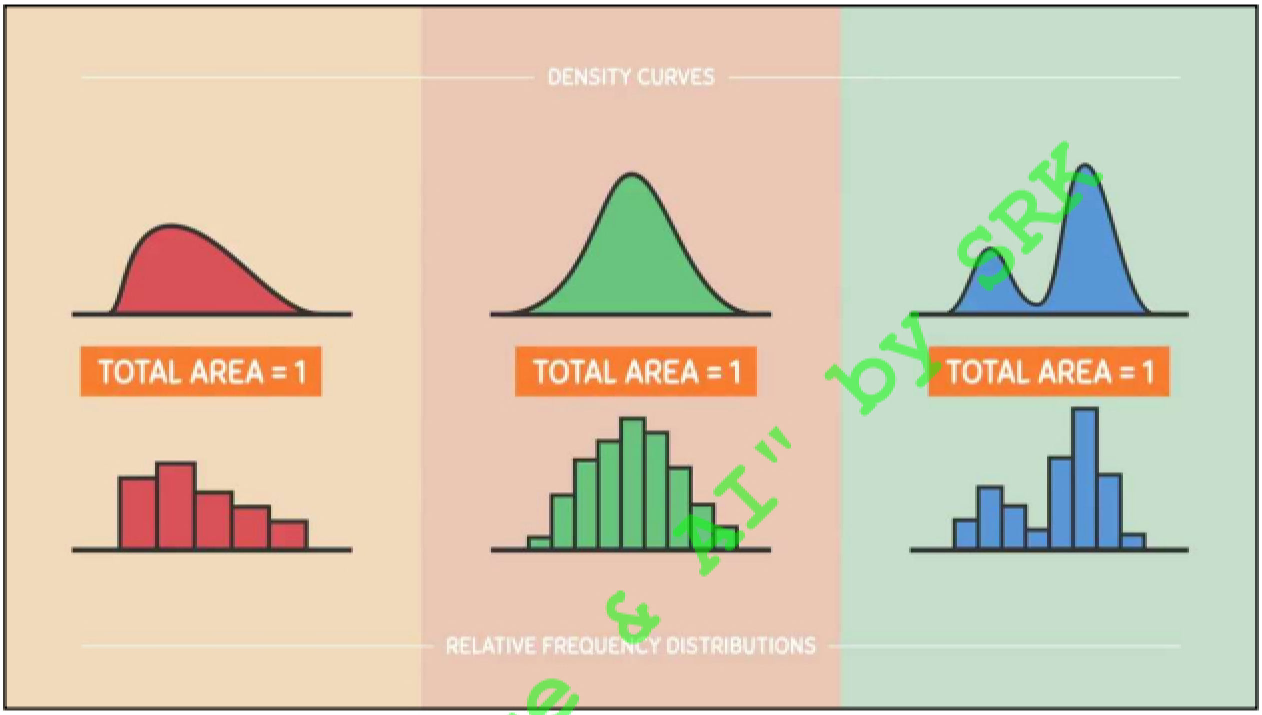


REVIEW

n = 20

"REGULAR" FREQUENCY DISTRIBUTION




DENSITY CURVES
TOTAL AREA = 1
TOTAL AREA = 1
TOTAL AREA = 1
RELATIVE FREQUENCY DISTRIBUTIONS

Probability

Variable :

- Chance of occurrence .
- Ex: rolling a die, tossing a coin

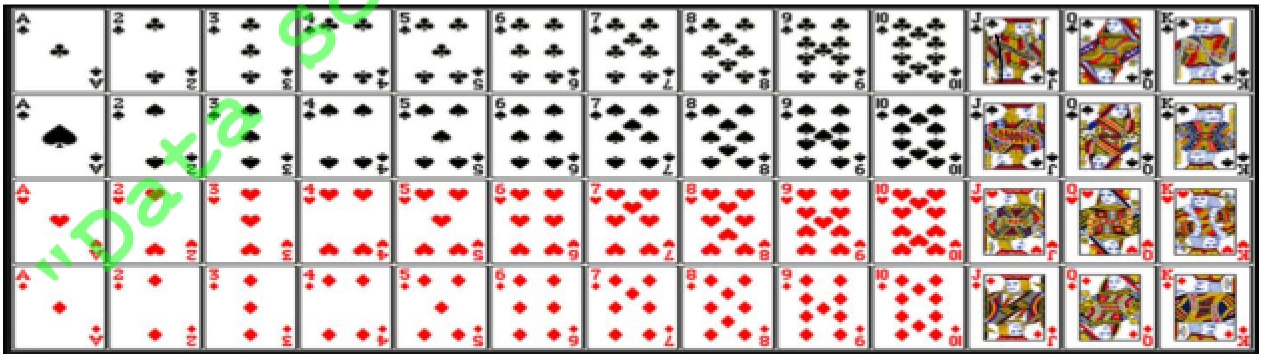
Random Variable:

A random variable is probability associated each possibility of variable .

It is a random because there is some chance associated with each possible value.

Probability = No. of interested events/total no. of outcomes

- Always probability value lies between 0 to1.
- Sum of all Probabilities =1



Suppose you have randomly picked a card from the card deck. What is the probability that this card will be?

- Bigger than 10?
- Equal to or Bigger than 10?
- Smaller than 3
- Greater than 4 and less than 8

If A & B are two independent events

P(A & B) = P(A) * P(B)

Ex: probability of getting Red & 9

P(A or B) = P(A) + P(B) – P(A & B)

Ex: probability of getting Red or 9

COVARIANCE

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

CORRELATON

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

where:

- cov is the covariance
- σ_X is the standard deviation of X
- σ_Y is the standard deviation of Y

CORRELATON

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

Range	Strength of association
0	No association
0 to ±0.25	Negligible association
±0.25 to ±0.50	Weak association
±0.50 to ±0.75	Moderate association
±0.75 to ±1	Very strong association
±1	Perfect association

OUTLIER

A DATA VALUE THAT IS NUMERICALLY
DISTANT FROM A DATA SET

OUTLIER

A DATA VALUE THAT IS NUMERICALLY
DISTANT FROM A DATA SET

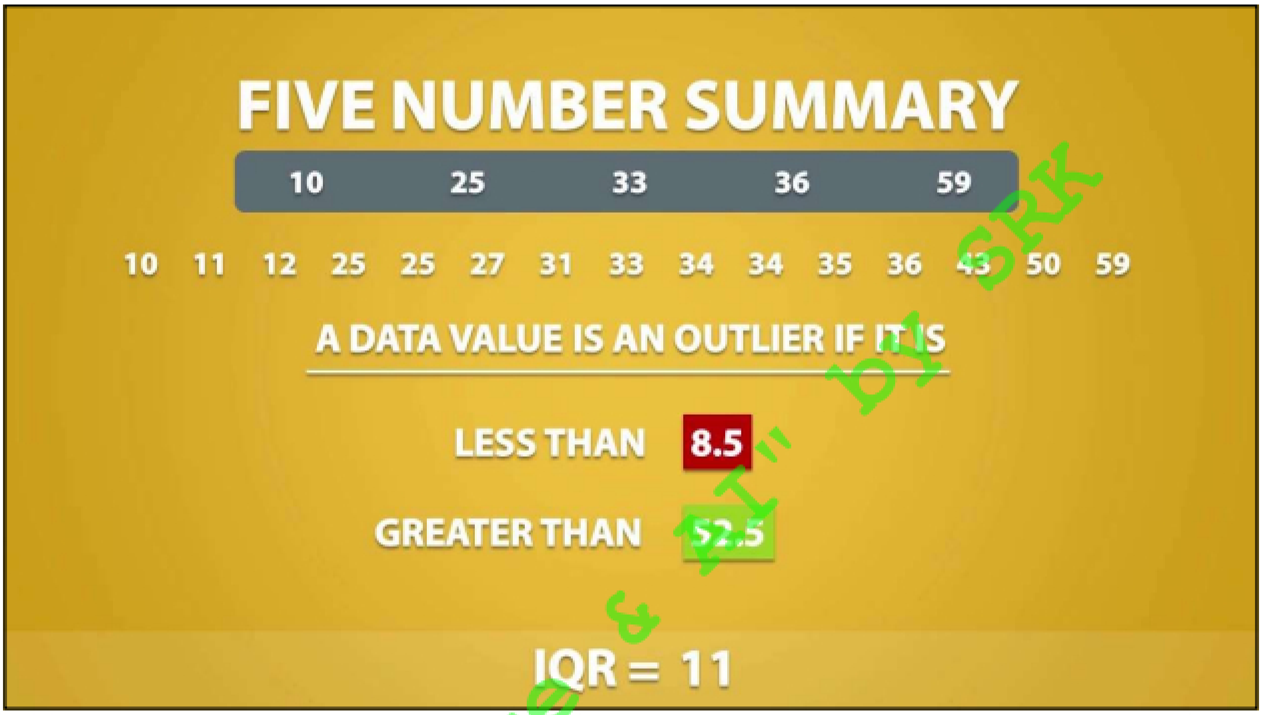
THE MEAN IS AFFECTED BY THE PRESENCE OF OUTLIERS

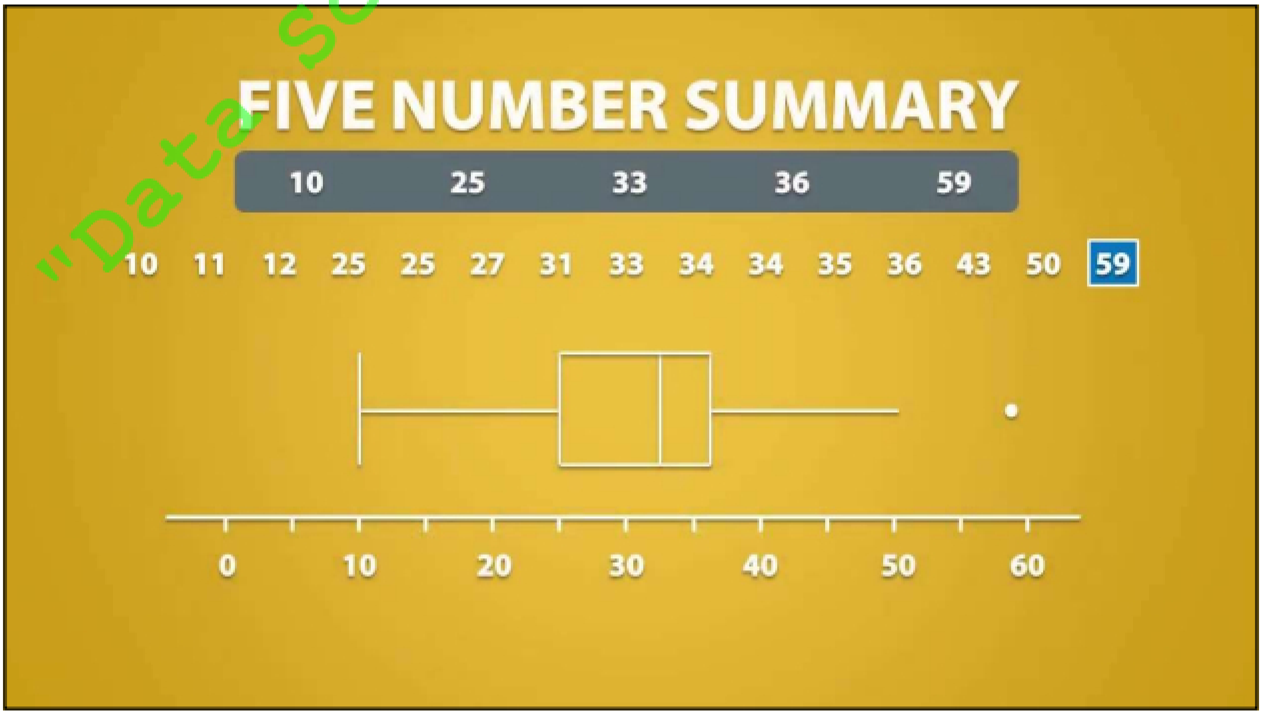
A DATA VALUE IS CONSIDERED TO BE AN OUTLIER IF..

DATA VALUE < Q1 – 1.5(IQR)

OR

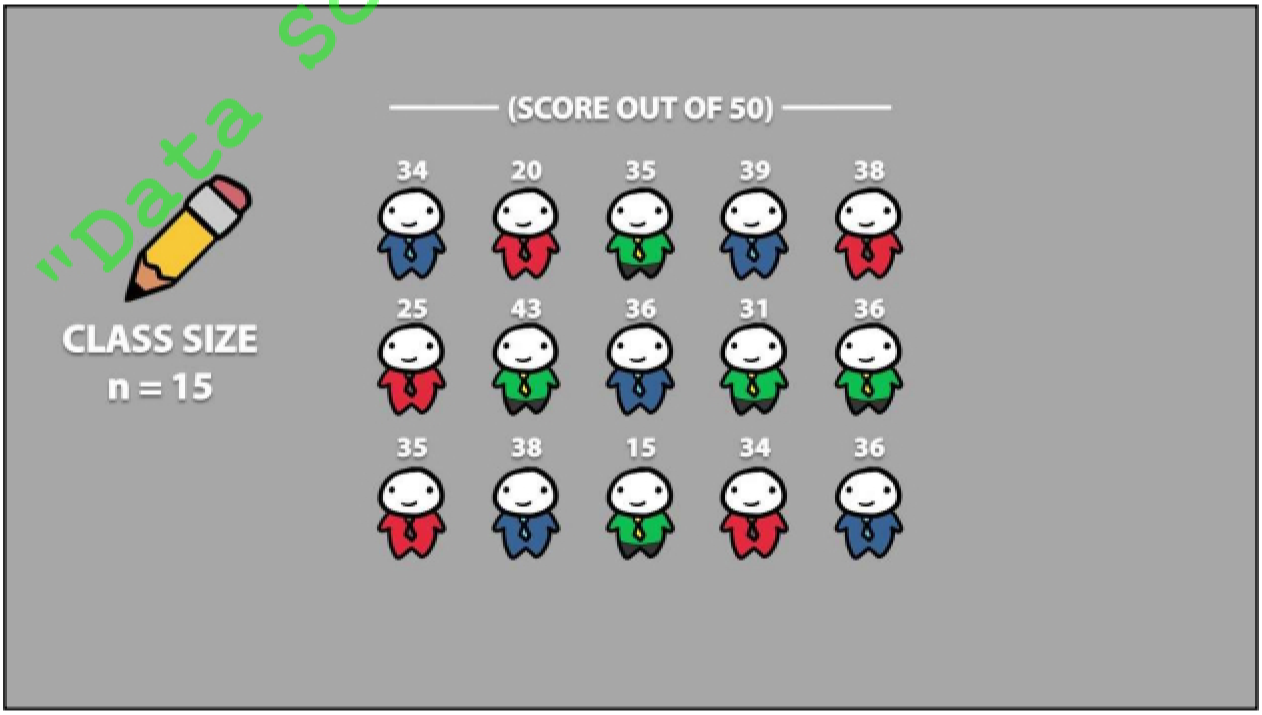
DATA VALUE > Q3 + 1.5(IQR)

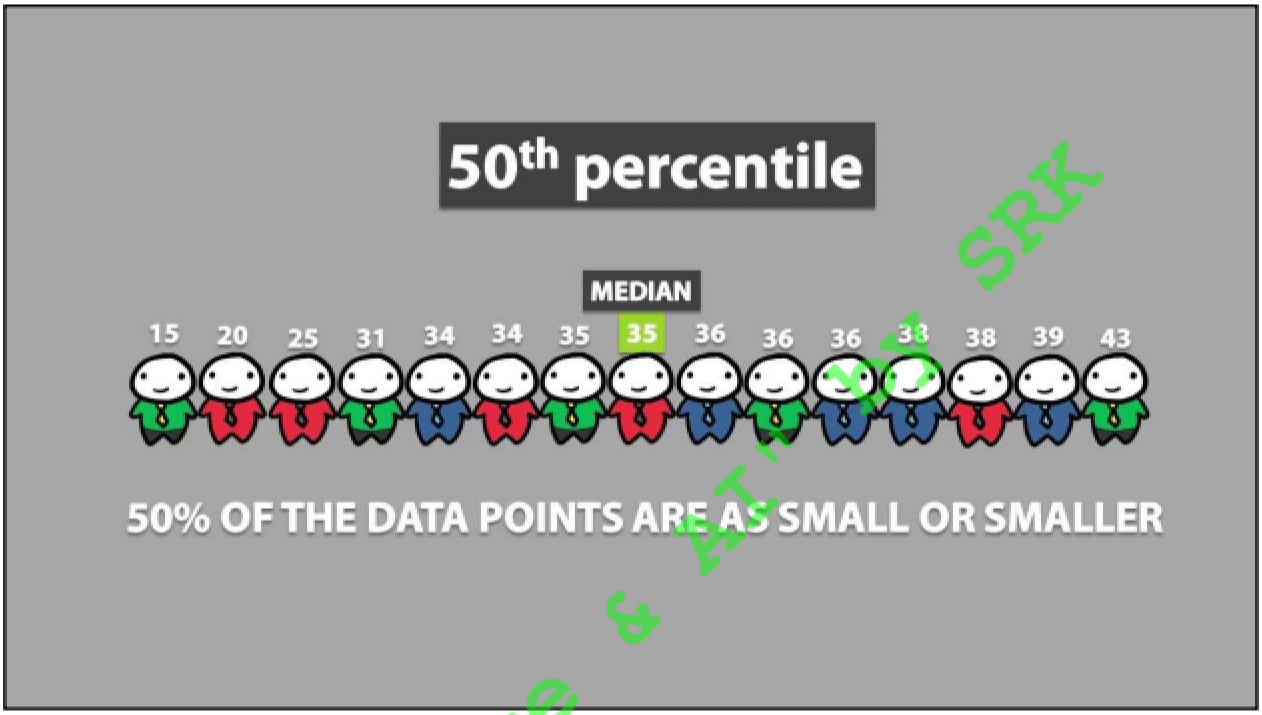

FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
A DATA VALUE IS AN OUTLIER IF IT IS
LESS THAN 8.5
GREATER THAN 52.5
IQR = 11


FIVE NUMBER SUMMARY
10 25 33 36 59
10 11 12 25 25 27 31 33 34 34 35 36 43 50 59
0 10 20 30 40 50 60

WHAT'S A PERCENTILE?

DESCRIBES THE PERCENTAGE OF DATA VALUES THAT FALL AT OR BELOW ANOTHER DATA VALUE





Percentiles: Computational Procedure

- Organize the data into an ascending ordered array
- Calculate the p th percentile location:

$$i = \frac{P}{100}(n)$$

- Determine the percentile's location and its value.
- If i is a whole number, the percentile is the average of the values at the i and (i+1) positions
- If i is not a whole number, the percentile is at the (i+1) position in the ordered array

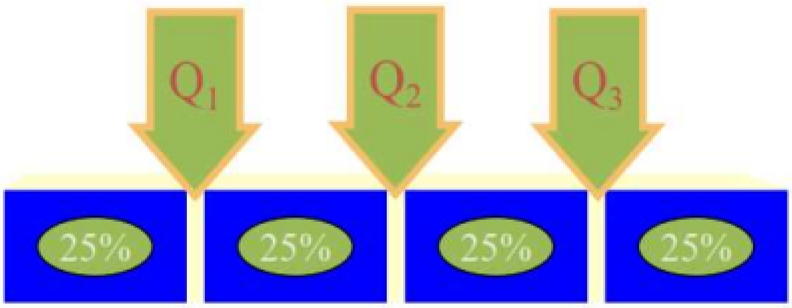
Percentiles: Example

- Raw Data: 14, 12, 19, 23, 5, 13, 28, 17
- Ordered Array: 5, 12, 13, 14, 17, 19, 23, 28
- Location of 30th percentile:

$$i = \frac{30}{100}(8) = 2.4$$

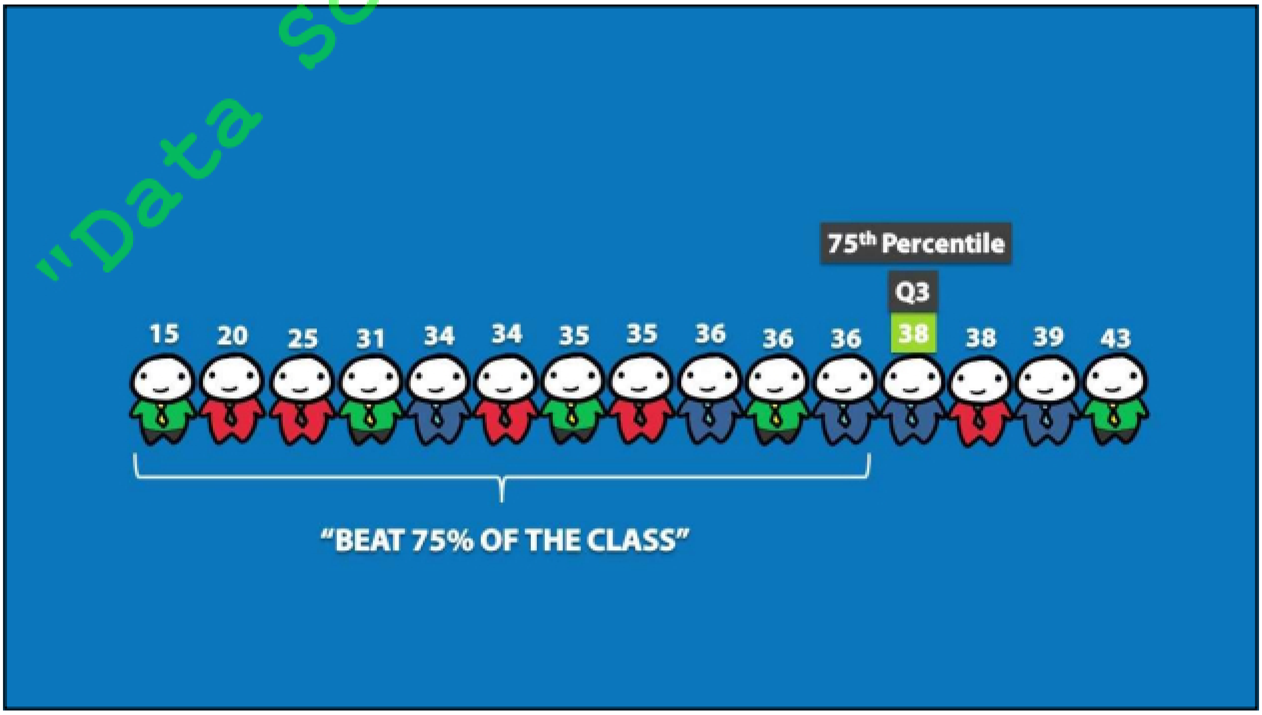
- The location index, i, is not a whole number; i+1 = 2.4+1=3.4; the whole number portion is 3; the 30th percentile is at the 3rd location of the array; the 30th percentile is 13.

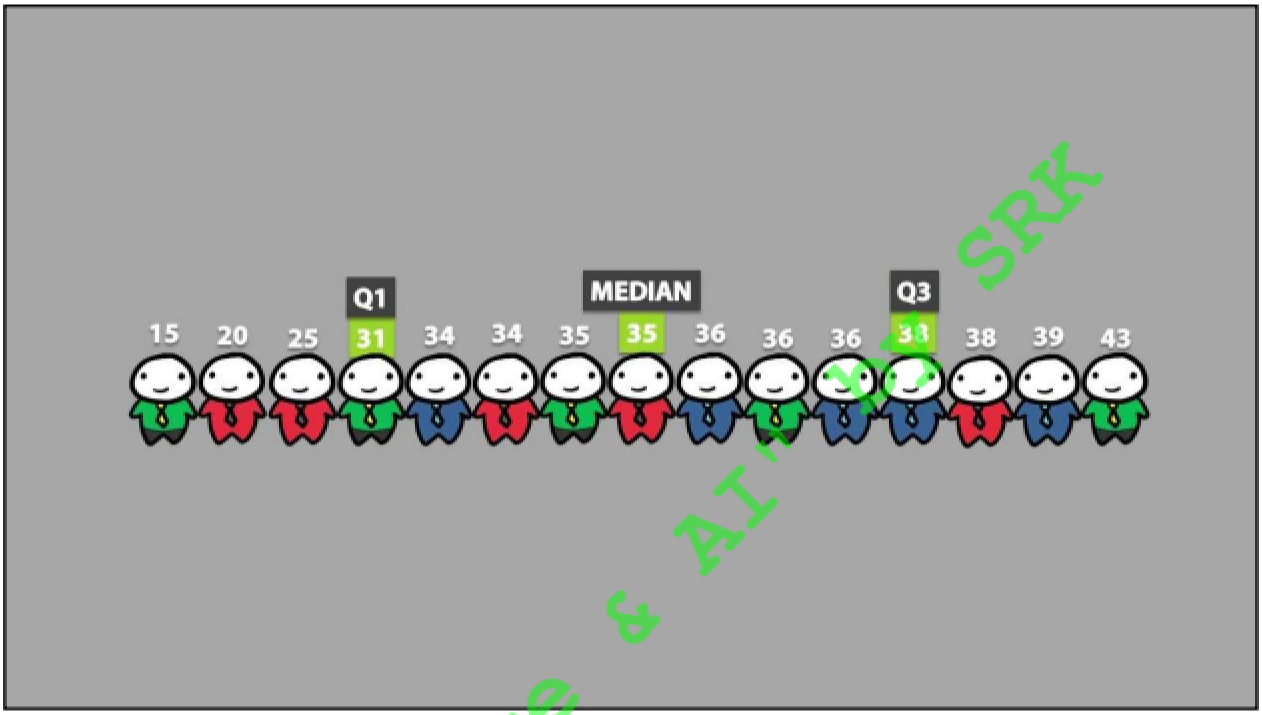
Quartiles



Quartiles

- Measures of central tendency that divide a group of data into four subgroups
- Q_1: 25% of the data set is below the first quartile
- Q_2: 50% of the data set is below the second quartile
- Q_3: 75% of the data set is below the third quartile
- Q_1 is equal to the 25th percentile
- Q_2 is located at 50th percentile and equals the median
- Q_3 is equal to the 75th percentile
- Quartile values are not necessarily members of the data set





Quartiles: Example

- Ordered array: 106, 109, 114, 116, 121, 122, 125, 129
- Q_1:

$$i = \frac{25}{100}(8) = 2 \qquad Q_1 = \frac{109+114}{2} = 111.5$$

- Q_2:

$$i = \frac{50}{100}(8) = 4 \qquad Q_2 = \frac{116+121}{2} = 118.5$$

- Q_3:

$$i = \frac{75}{100}(8) = 6 \qquad Q_3 = \frac{122+125}{2} = 123.5$$

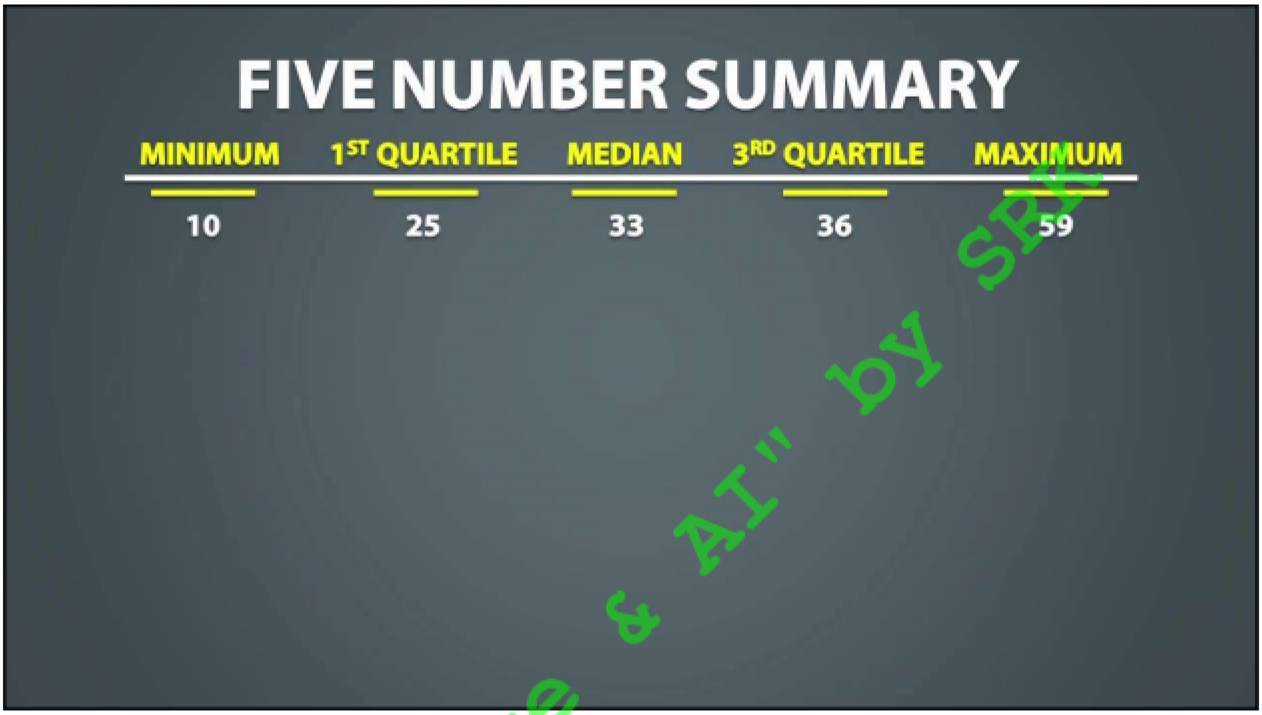
Interquartile Range

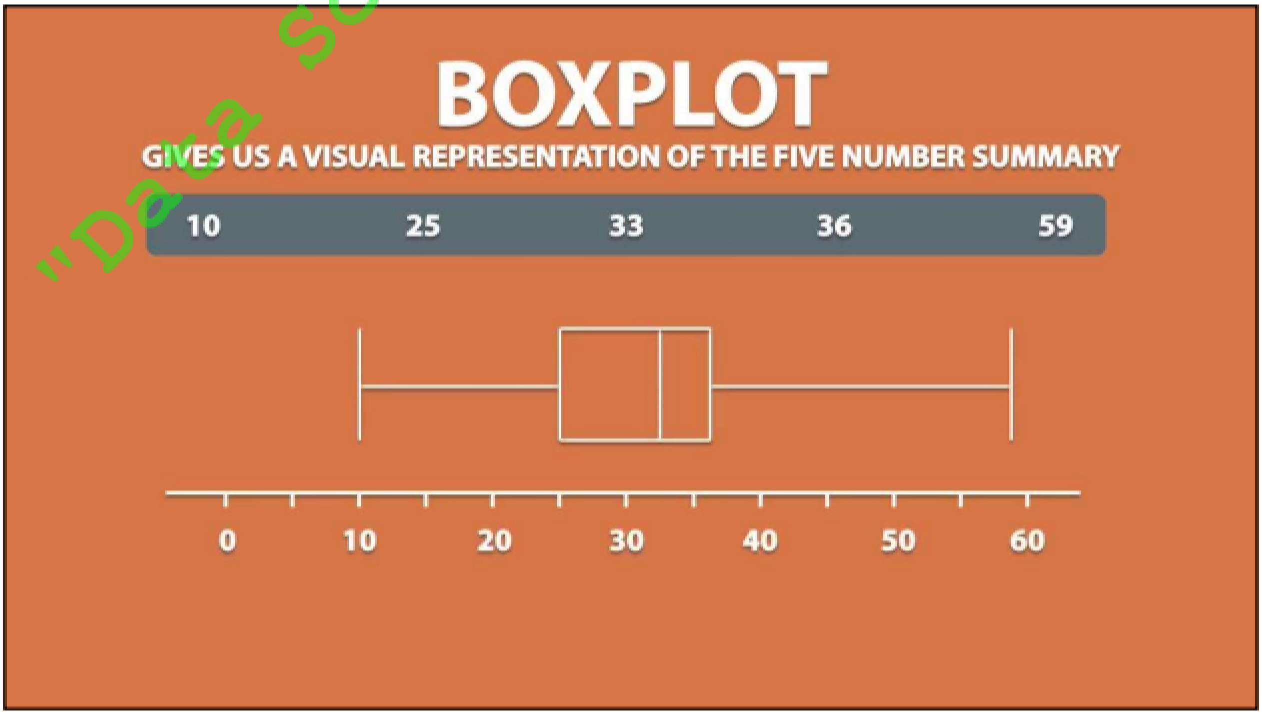
- Range of values between the first and third quartiles
- Range of the "middle half"
- Less influenced by extremes

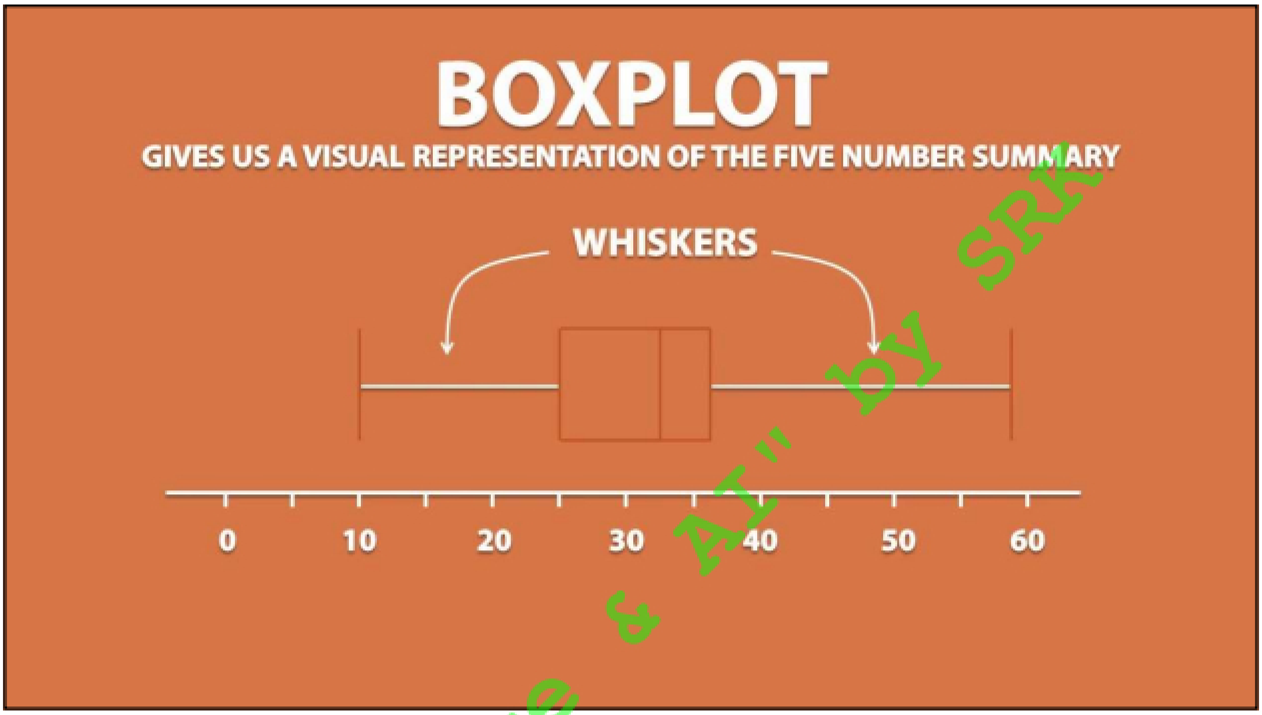
$$Interquartile\ Range = Q_3 - Q_1$$

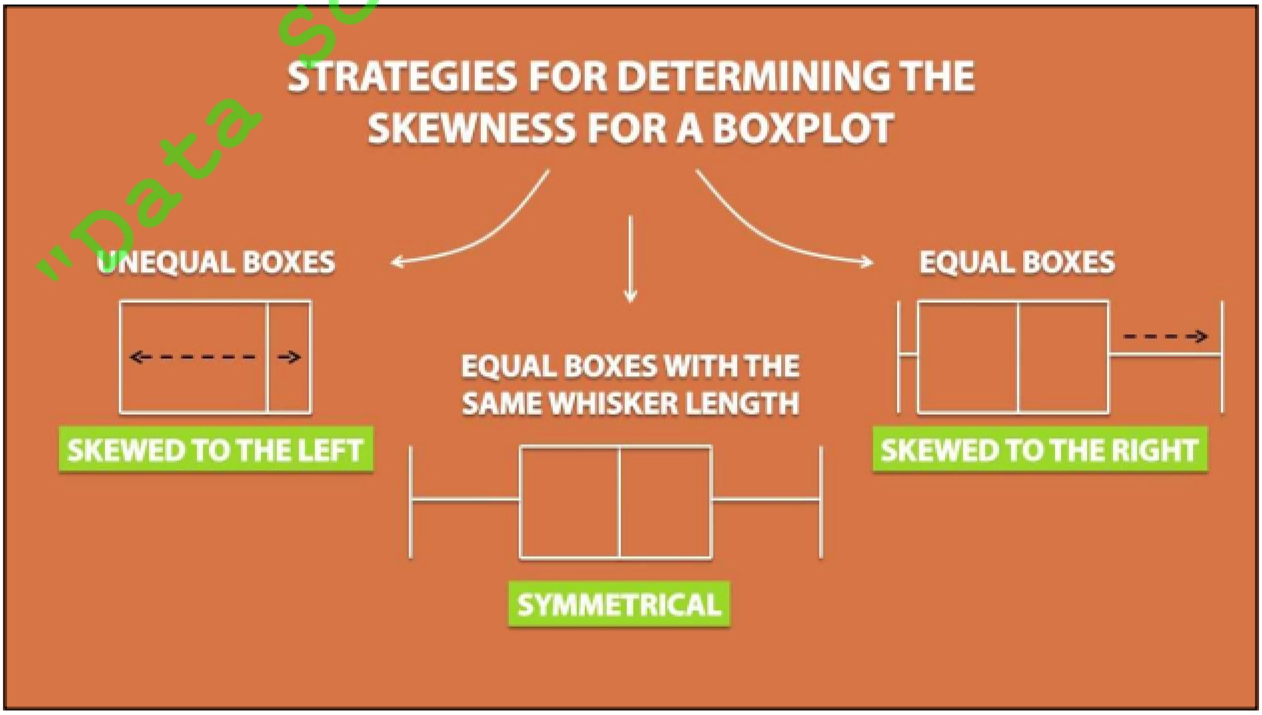
FIVE NUMBER SUMMARY

GIVES US A WAY TO DESCRIBE A DISTRIBUTION
USING ONLY FIVE NUMBERS


FIVE NUMBER SUMMARY
MINIMUM
1ST QUARTILE
MEDIAN
3RD QUARTILE
MAXIMUM
10
25
33
36
59


BOXPLOT
GIVES US A VISUAL REPRESENTATION OF THE FIVE NUMBER SUMMARY
10
25
33
36
59
0
10
20
30
40
50
60


BOXPLOT
GIVES US A VISUAL REPRESENTATION OF THE FIVE NUMBER SUMMARY
WHISKERS
0
10
20
30
40
50
60


STRATEGIES FOR DETERMINING THE
SKEWNESS FOR A BOXPLOT
UNEQUAL BOXES
SKEWED TO THE LEFT
EQUAL BOXES WITH THE
SAME WHISKER LENGTH
SYMMETRICAL
EQUAL BOXES
SKEWED TO THE RIGHT